DOCUMENT RESUME

ED 219 403 TM 820 414

AUTHOR Benson, Jeri
TITLE Detecting Item Bias in Affective Scales.
PUB DATE Mar 82
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Affective Measures; Blacks; *Ethnic Groups; *Factor Analysis; Grade 8; Hispanic Americans; *Item Analysis; Junior High Schools; *Self Concept Measures; *Test Bias; Whites
IDENTIFIERS Confirmatory Factor Analysis; *LISREL Computer Program

ABSTRACT

The purpose of this study was to demonstrate the use of confirmatory factor analysis to investigate the existence of item bias in an affective scale measuring self concept. The LISREL IV program was used to detect test item bias and factorial invariance. The ethnic groups compared in the study were Black, Hispanic, and White. The obtained factor structure was compared statistically across the three groups. The results of the study indicated a need to consider affective instruments, as well as cognitive instruments, in the study of item bias. If bias were found to exist in an affective scale, recognition of the effects of that bias and the influence on the interpretation of the resulting score would be important. LISREL analysis provided a method to investigate differential response patterns from affective scales simultaneously that could not be objectively analyzed by classical factor analysis. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED219403

Detecting Item Bias in Affective Scales

Jeri Benson

University of Southern California

The purpose of the study was to demonstrate the use of confirmatory factor analysis to investigate the existence of item bias in an affective scale measuring self-concept. The data were collected from three ethnic groups (Black, Hispanic and White) in grade eight. The factor structure across ethnic groups was compared simultaneously using LISREL analyses to determine whether the structures were statistically different.

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Benson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

A paper presented at the annual meeting of
the American Educational Research Association
New York, March 1982

TM 820 414

Detecting Item Bias in Affectives Scales

Jeri Benson

University of Southern California

Much of the research to date on studying test-item bias has been focused on cognitive instruments. This research is certainly warranted in that most of the testing that takes place is in the cognitive domain. In education, achievement and aptitude test results are constantly used to make not only decisions about individuals, but also about curricular and school effectiveness. However, researchers are increasingly using affective variables as educational outcomes themselves or as moderators of achievement that might impact upon cognitive test results. Thus, there appears to be a need to consider whether and to what extent item bias may effect the results of affective scales. Consequently, the interest in studying item bias may need to be extended to include the affective domain.

The operational definition most often used for deciding whether test or item bias in a cognitive instrument is present is when ..."equally able individuals, from different groups, do not have equal probabilities of success" (Shepard, Camilli & Averill, 1981, p. 319). Based upon the above definition, several empirical methods of identifying item bias in cognitive tests have been developed (Angoff, 1972; Merz, 1973; Rudner, 1977; Scheuneman, 1975, 1979; Wright, Mead & Draba, 1976) as well as empirical studies

comparing the effectiveness of the various techniques (Ironson & Subkoviak, 1979; Rudner, Getson & Knight, 1980; Scheuneman, 1980; Shepard et al., 1981). All of the item detection methods developed to date for cognitive instruments require knowledge of a correct or keyed answer in order to identify biased items. With affective instruments correct or keyed responses are not relevent as one is usually more interested in the intensity of the attitude and not the 'correct' attitude. Therefore, the methods developed to date for detecting item bias in cognitive instruments are not appropriate for a study of item bias in affective instruments.

Green (1975) noted that the issue of bias was basically a question of validity, specifically of differential validity between two or more groups, generally a minority was compared to the population. He stated that test bias can occur when the "test description is more erroneous for one group than another" (p. 16). That is, the scores do not mean the same thing for each group. Green stated that if the factor structure for a cognitive instrument differs for the groups taking the test then that may be an indication of instrument bias. His assumption has been extended to the affective domain in the present study where biased items will be thought to exist if the factor structure varies across groups.

Green (1975) discussed bias as it related to three different test uses: prediction, selection and description. In the present study the researcher was primarily concerned with bias as it related to the descriptive use of a test.

For descriptive affective instruments, both content and construct validity is necessary. It is quite possible for an instrument to possess content validity; however, the items may be responded to differentially by members of each group and thus the construct underlying the instrument may differ for the groups. When this occurs the items can be thought to elicit different response patterns across groups and will represent an operational definition of bias for an affective scale. Therefore, the measurement, if biased, would result in tests that measure different things for different groups with the same items. Any comparisons between groups or conclusions about groups based upon items that measure differentially in this manner would be inappropriate.

In order to assess the construct validity of an affective measure and to detect biased items one must go to the internal structure of the instrument (Angoff & Ford, 1973; Cardell & Coffman, 1964; Cleary & Hilton, 1968). Factor analysis of the items has been used to identify the differential score variance across groups (Green & Draper, 1972; Merz, 1973). The purpose of the present study was to determine whether an affective scale would describe the attitudes of students from three ethnic groups in a similar fashion. Would the three ethnic groups respond to the same items in the same manner so that an invariant factor structure based upon the three groups would be found to underly the scale? Exploratory factor analysis can provide an indication of how similar the factor structure for each ethnic group was and

also a means to examine how the items functioned for each ethnic group. However, exploratory factor analysis alone does not provide objective data on the similarity of factor structures. Thus, a confirmatory factor analysis procedure suggested by McGaw and Joreskog (1971) using the LISREL IV program to test simultaneously for factorial invariance across the three groups is proposed for the detection of test-item bias in a affective scale.

METHOD

Sample

The population consisted of 29,840 eighth and 34,492 fifth grade students who responded to an affective survey that was administered at the end of the year as part of a court-ordered integration study. For purposes of the present study, only the eighth grade student responses were considered. A four percent random sample from the population of eighth grade responses was used (N=1264). The ethnic breakdown of the sample was Hispanic (n=377), White (n=492) and Black (n=395).

Instrument

The instrument for use in the study was one originally developed by ETS for a nation-wide integration study (Forhand, Ragosta & Rock, 1976). The instrument was further re-

fined by adding or changing items to meet local needs. The eighth grade questionnaire contained 32 items to assess self concept and racial attitudes.

Procedures

The procedure being proposed for investigating item bias in affective scales is two-fold. First, a classical factor analysis with Varimax rotation was run on one ethnic group (the Hispanic sample was randomly chosen) to determine the factor structure for that group. This step was included to find out roughly how many factors were underlying the 32 items and also which items loaded highest on which factors in order to use only the most highly correlated items for the LISREL analysis. Second, the obtained factor structure was then compared statistically using the LISREL IV program developed by Jöreskog and Sörbom (1978) across the three groups. The covariance matrix was used as input for the LISREL program. Since there is a large number of models that can be tested using the LISREL procedure, a systematic approach was used. Given the factor structure that was obtained from the classical factor analysis three questions were of interest. Did the same number of factors underly the three groups? Were the correlations between factors the same across the three groups? Were the errors of measurement the same across the three groups? From these questions four models thought to represent the data were tested.

a) Model 1 - Each group has the same factor structure and the same errors of measurement

b) Model 2 - Each group has a different factor structure but the same errors of measurement

c) Model 3 - Each group has the same factor structure but different errors of measurement

d) Model 4 - Each group has a different factor structure and different errors of measurement

The correlations among factors for the above models was not constrained, thus allowing for possible differences within each group to emerge.

Model testing using the LISREL technique has been used previously to study structural invariance of an instrument among several populations (Jöreskog, 1971; McGaw & Jöreskog, 1971; Sörbom, 1974). If the comparisons of the factor structures for the three groups were found to be statistically different then that would be an indication of bias. Individual factors for each ethnic group would then be investigated to determine which items led to the differential results. In this case bias was not considered to have a negative connation in which one group may be unjustly treated based upon their scores, but merely an indication that group differences would need to be considered when scores were interpreted.

The LISREL program requires that three matrices be specified prior to analysis: the lambda pattern matrix, the psi matrix and the theta epsilon matrix. The lambda pattern matrix allows the user to identify which variables are free and thought to load on a given factor and which variables are fixed to zero and thought not to load on a factor. In

this way, the hypothesized factor pattern based upon the above four models can be tested. The psi matrix allows one to see if the factors are correlated by setting them free or uncorrelated by setting them to be fixed at zero. Finally, the theta epsilon matrix which provides an estimate of the error or uniqueness of each variable must be specified. In addition, when a variable was thought to be invariant it had to be specified at the beginning of the program prior to establishing the lambda pattern matrix and the psi and theta epsilon matrices. For example, in Model 1 where both the factor structure and the errors were tested for invariance, these variables were analyzed simultaneously by specifying the invariant mode for both the lambda (factor) matrix and theta (error) matrix. For Model 2 only the theta matrix was held invariant across groups and for Model 3 only the lambda matrix was held invariant across groups. Model 4 was analyzed by setting all variables free to obtain an unrestricted solution for both factors and errors. Models 2 and 4 if significantly different from Models 1 and 3 respectively would be an indication of bias. However, Model 4 would represent the strongest case for differential responding and the greatest evidence of item bias in the scale.

Statistical significant between factor structures is obtained from the LISREL program by the chi-square statistic. In additon, estimates of the factor loadings and item uniquesnesses are provided to indicate model-data fit.

RESULTS and DISCUSSION

From the classical factor analysis of the Hispanic sample a three factor solution (based upon the 32 items) seemed to best represent the data in terms of factor importance and also psychological meaningfulness. The three factors were composed of self-concept and locus of control items and were labeled liking for school (8 items), self-rated ability (6 items) and internality (5 items).

Thus, a separate covariance matrix for each ethnic group composed of the 19 items representing the above three factors were used as input to the LISREL IV program in order to obtain a simultaneous factor solution across the three groups of students. The chi-square test of model-data fit is reported in Table 1 for the four models that were tested. The lower the chi-square value the better the model-data fit. Thus, a non-significant chi-square value is desired as it would indicate no differences between the model proposed and the data (Bentler & Bonett, 1980). However, for large sample sizes, this result is extremely unlikely.

Insert Table 1 about here

Model 1, where an invariant factor structure and errors of measurement was fit to the three groups, a significant chi-square (χ^2= 1172.3167 with 523 degrees of freedom p <.001) was revealed. The significant result indicated that the covariance matrix generated by the three factor solution

was significantly different from the original covariance matrix used as input. Likewise Models 2 - 4 revealed statistically significant chi-square values at $p < .001$ (see Table 1). Since the statistical significance of the chi-square test can be influenced by sample size, Bentler and Bonett (1980) recommend the comparison of alternate models to determine the relative effectiveness of one model over another in terms of fit. Bentler and Bonett suggest the following chi-square difference test be used to test the significance of model data fit:

$$\chi^2_{1-2} = \chi^2_1 - \chi^2_2 \text{ and degrees of freedom} = df_1 - df_2, \quad (1)$$

where χ^2_1 represented one hypothesized model and χ^2_2 represented an alternate hypothesized model with their corresponding difference in degrees of freedom. If the difference chi-square (χ^2_{1-2}) is statistically significant then the alternate model fit the data better than the first hypothesized model.

Since four models were hypothesized in the present study, formula 1 can be applied to make comparisons between each of the observed chi-square values in Table 1 to determine which of the four models best represented the data. When Models 1 and 2 were compared the result ($\chi^2_{1-2} = 91.87$ and 38 degrees of freedom) was statistially significant at $p < .001$. Thus, Model 2, which represented a different factor structure and same errors was a better fit to the data than Model 1, which represented an invariant factor structure and invariant errors across the three groups. The other model comparisons that were found to be statistically different

were: Model 1 - Model 3 (χ^2_{1-3} 93.98 and 38 degrees of freedom, p <.001), Model 1 - Model 4 (χ^2_{1-4} 192.39 and 76 degrees of freedom, <.001), Model 2 - Model 4 (χ^2_{2-4} 100.52 and 38 degrees of freedom, p <.001) and Model 3 - Model 4 (χ^2_{3-4} = 98.41 and 38 degrees of freedom, p <.001). The comparison between Model 2 and Model 3 could not be made since the degrees of freedom was the same for the two models. From this analysis, Model 4 containing the different factor structure and different errors appeared to be a better fit of the data than the other models.

A further test to determine incremental model-data fit was suggested by Bentler and Bonett (1980). An incremental index of fit provides information regarding the practical significance of the alternative models independent of sample size. The fit index is defined as follows:

$$\Delta_{12} = \chi^2_1 - \chi^2_2 \; / \; \chi^2_1 \quad , \qquad (2)$$

where χ^2_1 and χ^2_2 represent the two proposed models thought to underlie the data. Bentler and Bonett suggest that the most restrictive model be used in the denominator of Delta. Since Model 1 was thought to be the most restrictive (factor structure and errors invariant across the three groups) it was placed in the demoninator. The Delta coefficients for the model comparisons have been reported in Table 2. The values of Delta range between zero and one.

Insert Table 2 About Here

From the results provided in Table 2, it appeared that Model 4 (different factor structure - different errors) provided the best increment of fit when compared to Model 1 (same factor structure - same errors).

At this point sufficient evidence was indicated to conclude that the three ethnic groups were perhaps responding to the same items in a differential manner and an inspection of the three separate factor structures from the LISREL analysis would be in order. From the separate factor structures one was able to identify items the students responded to differentially. On Table 3 is shown the factor loadings (lambda) and error (theta) estimates for the White sample. On Table 4, the factor intercorrelations (psi matrix) for the White sample is presented.

Insert Tables 3 and 4 about here

Each of the item-factor loadings are statistically significant at $p < .05$ (execpt for item 34). The observed theta values are provided as an indication of how much error is in an item. Since the covariance matrix was used as input the theta values are relative as the total variance for an item can be greater than one. From Table 4, the intercorrelations of the three factors for the White group indicate that Factors 1 and 2 (liking for school and self-rated ability) are highly correlated and Factors 1 and 3 (liking for school and internality) were highly correlated. The correlation between Factors 2 and 3 (self-rated ability and internality) was not significant.

Tables 5 and 6 provide the factor loadings, item error estimates and factor intercorrelations for the Hispanic sample.

Insert Tables 5 and 6 about here

Each of the item factor loadings were significant at $p < .05$ (except for items 12 and 28). The factor intercorrelations (from Table 6) show a pattern quite different from the White sample. Factors 1 and 2 are correlated, but not as highly as in the White sample. Factors 1 and 3 were negatively correlated, whereas in the White sample these factors were highly correlated in a positive direction. Therefore, the Hispanic students did not respond to the same set of items in the same pattern as the White students. For example, when the Hispanic students indicated a high rating on liking for school, they indicated a corresponding low rating on internality items. This finding may be due in part to cultural differences between Hispanic and White students and as such the items elicited differential response patterns as noted in the data. The pattern of correlations between Factors 2 and 3 for the Hispanic sample is similar to the pattern for the White sample (no correlation).

The results of the factor analysis for the Black sample is shown on Table 7 and the factor intercorrelations are reported on Table 8.

Put Tables 7 and 8 about here

All of the factor loadings were significant at $p < .05$ (except items 36 and 20). The factor intercorrelations for the Black sample follow a pattern similar to the White group in direction, but not in magnitude. Factors 1 and 2 and 1 and 3 are positively correlated, but not to the same degree as the White sample. Again Factors 2 and 3 were not shown to be related.

Items from each of the three factors are reported in Table 9 to add insight to the results presented. However, an analysis of why the differential response pattern occured for the items was not the focus of the present paper.

Put Table 9 about here

From the results presented, a rationale for studying bias in affective scales seems warranted. The three ethnic groups did not respond to the same items in the same manner as noted by the lower chi-square value for Model 4 (factors structure and errors different across groups) and the fact that the increment of fit index indicated Model 4 best represented the original data set. When the separate factor structures were compared it was observed that different items failed to load significantly on the three factors and the factor intercorrleations across groups differed for the Hispanic and White groups for Factors 2 and 3. Why these

items were responded to in a differential manner needs to be investigated in a subsequent paper. The instrument in the present paper was used only to explore the possiblity of LISREL procedures as being a useful technique to detect bias in an affective scale. Subsequent research efforts should also consider whether a three factor solution is the best representation of the data at hand since all of the models revealed significant chi-square values. On this point however, Jöreskog (1971) has stated that even if the is large there still may be reasons to consider the model since there will always be discrepancies between the actual covariance matrix and the model postulated. The decison to accept a model is not always based upon statistics, but upon the usefulnes it provides (p. 421).

Since psychometricians as well as the general public are well aware of item bias and its effects on cognitive instruments the results of the present study indicate a need to also consider affective instruments in the study of item bias. The scale under study was shown to be differentially valid for separate ethnic groups and that while certain items may be content valid for the task, students do not respond to those items in the same manner. This research represents an initial attempt in applying confirmatory factor analysis in the detection of bias in affective scales. If bias is found to exist in an affective scale then it becomes important to recognize how the bias may effect the interpretation of the resulting score. It may be that separate factor scores for each ethnic group would need to be

derived before a score could be interpreted if bias were present. In conclusion, LISREL analysis can be thought to provide a means for investigating differential response patterns across different groups simultaneously from affective scales that heretofore could not be objectively analyzed by classical factor analysis. Thus, LISREL procedures seem to appear to be a viable technique for investigating bias in affective scales as defined by the differntial factorial composition of the scale.

References

Angoff, W. A technique for the investigation of cultural differences. Paper presented at the annual meeting of the American Psychological Association, Honolulu, 1972.

Angoff, W.H. & Ford, S.F. Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 1973 10, 95-105.

Bentler, P.M. & Bonett, D.G. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 1980, 88, 588-606.

Cardell, C & Coffman, W.R. A method for comparing performance on different groups on the items of a test. Research Memorandum 64-61, Princeton: Educational Testing Service, 1964.

Cleary, T.A. & Hilton, T.L. An investigation into item bias. Educational and Psychological Measurement, 1968, 8, 61-75.

Forehand, G.A., Ragosta, M. & Rock, D.A. Conditions and processes of effective school integration. (U.S. Department of Health, Education and Welfare) Washington, D.C.: Office of Planning, Budgeting and Evaluation, 1976. (ERIC No. 131 155)

Green, D.R. What does it mean to say a test is biased? Education and Urban Society, 1975, 8, 33-52.

Green, D.R. & Draper, J.F. Exporatory studies of bias in achievement tests. Paper presented at the annual meeting of the American Psychological Association, Honolulu, 1972.

Ironson, G.H. & Subkoviak, M.J. A compairson of several methods of assesing item bias. Journal of Educational Measurement, 1979, 16, 209-255.

Jöreskog, K.G. Simultaneous factor analysis is several populations. Psychometrika, 1971 36, 409-426.

Jöreskog, K.G. & Sörbom, D. LISREL IV: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: International Educational Services, 1978.

McGaw, B. & Jöreskog, K.G. Factorial invariance of ability measures in groups differing in intelligence and socioeconomic status. British Journal of Mathematical and Statistical Psychology, 1971, 24, 154-168.

Merz, W.R. Factor analysis as a technique in analyzing item bias. Paper presented at the annual meeting of the California Educational Research Association, Los Angeles, 1973.

Rudner, L.M. An approach to biased item identification using latent trait measurement theory. Paper presented at the annual meeting of the American Educational Research Association, New York, 1977.

Rudner, L.M., Getson, R.R. & Knight, D.L. A monte carlo comparison of seven biased item detection techniques. Journal of Educational Measurement, 1980, 17, 1-10.

Scheuneman, J. A new method of assessing bias in test items. A paper presented at the annual meeting of the American Educational Research Association, Washington, 1975.

Scheuneman, J. A method of assessing bias in test items. Journal of Educational Measurement, 1979, 16, 143-152.

Scheuneman, J. Consistency across administration of certain indices of bias in test items. A paper presented at the annual meeting of the American Educational Research Association, Boston, 1980.

Shepard, L., Camilli, G. & Averill, M. Comparison of procedures for detecting item bias with both internal and external ability criteria. Journal of Educational Statistics, 1981, 6, 317-375.

Sörbom, D. A general method for studying differences in factor means and factor structures between groups. British Journal of Mathematical and Statistical Psychology, 1974, 27, 229-239.

Wright, B.D., Mead, R.J. & Draba, R. Detecting and correcting test item bias with a logistic response model. Research Memorandum 22, Statistical Laboratory, Department of Education, University of Chicago, 1976.

Table 1

Models Tested Regarding the Factor Structure and Error of Measurement across the three Groups

Model	Description	χ^2	df
1	Same Factor Structure Same Errors	1172.3167	523
2	Different Factor Structure Same Errors	1080.4436	485
3	Same Factor Structure Different Errors	1078.3390	485
4	Different Factor Structure Different Errors	979.9282	447

Table 2

Delta Coefficients[a] For Model Comparisons

	Model 1	Model 2	Model 3	Model 4
Model 1	--	.078	.080	.164
Model 2		--	.002	.093
Model 3			--	.091
Model 4				--

[a] Delta Coefficients range from 0 to 1.

Table 3 [a]

Factor Loadings (Lambda) and Error (Theta)
Estimates for the White Sample

N = 433

Item	Factor 1	Factor 2	Factor 3	Theta
5	.099	0.0	0.0	.091
6	.513	0.0	0.0	1.150
13	.243	0.0	0.0	.185
23	.302	0.0	0.0	.555
25	.270	0.0	0.0	.325
27	.230	0.0	0.0	.181
29	.303	0.0	0.0	.138
39	.323	0.0	0.0	.321
21	0.0	.341	0.0	.380
30	0.0	.381	0.0	.256
34	0.0	.076	0.0	.427
36	0.0	.169	0.0	.684
37	0.0	.213	0.0	.645
38	0.0	.275	0.0	.470
12	0.0	0.0	.233	.194
18	0.0	0.0	.268	.502
20	0.0	0.0	.342	.670
22	0.0	0.0	.320	.694
28	0.0	0.0	.145	.229

Table 4 [b]

Factor Intercorrelations (Psi Matrix)
For the White Sample

	Factor 1	Factor 2	Factor 3
Factor 1	1.000		
Factor 2	.662	1.000	
Factor 3	.617	.099	1.000

a,b Factor 1 = liking for school
Factor 2 = self rated ability
Factor 3 = internality

Table 5 [a]

Factor Loadings (Lambda) and Error (Theta)
Estimates for the Hispanic Sample

N = 343

Item	Lambda Factor 1	Lambda Factor 2	Lambda Factor 3	Theta
5	.085	0.0	0.0	.057
6	.434	0.0	0.0	1.142
13	.255	0.0	0.0	.168
23	.378	0.0	0.0	.427
25	.270	0.0	0.0	.381
27	.276	0.0	0.0	.168
29	.271	0.0	0.0	.110
39	.317	0.0	0.0	.335
21	0.0	.314	0.0	.423
30	0.0	.272	0.0	.319
34	0.0	.242	0.0	.366
36	0.0	.420	0.0	.516
37	0.0	.261	0.0	.598
38	0.0	.361	0.0	.357
12	0.0	0.0	-.017	.230
18	0.0	0.0	.313	.548
20	0.0	0.0	.520	.360
22	0.0	0.0	.293	.635
28	0.0	0.0	.034	.245

Table 6 [b]

Factor Intercorrelations (Psi Matrix)
for the Hispanic Sample

	Factor 1	Factor 2	Factor 3
Factor 1	1.000		
Factor 2	.336	1.000	
Factor 3	-.339	-.064	1.000

[a,b] Factor 1 = liking for school
Factor 2 = self rated ability
Factor 3 = internality

Table 7 [a]

Factor Loadings (Lambda) and Error (Theta)
Estimates for the Black Sample

N = 337

Item	Factor 1	Factor 2	Factor 3	Theta
5	.034	0.0	0.0	.047
6	.316	0.0	0.0	1.601
13	.209	0.0	0.0	.206
23	.325	0.0	0.0	.553
25	.216	0.0	0.0	.375
27	.276	0.0	0.0	.173
29	.226	0.0	0.0	.144
39	.267	0.0	0.0	.331
21	0.0	.355	0.0	.326
30	0.0	.289	0.0	.320
34	0.0	.249	0.0	.315
36	0.0	.078	0.0	.710
37	0.0	.125	0.0	.611
38	0.0	.285	0.0	.476
12	0.0	0.0	.193	.211
18	0.0	0.0	.289	.605
20	0.0	0.0	.033	.724
22	0.0	0.0	.334	.640
28	0.0	0.0	.115	.236

Table 8 [b]

Factor Intercorrelations (Psi Matrix)
For the Black Sample

	Factor 1	Factor 2	Factor 3
Factor 1	1.000		
Factor 2	.425	1.000	
Factor 3	.534	-.022	1.000

[a,b]

Factor 1 = liking for school

Factor 2 = self rated ability

Factor 3 = internality

Table 9

Items Representing Factors 1-3

Factor 1 (Liking for School)

5) Think about most of the work you have to do in school. Is it too hard, too easy, or just about right?

6) How do your parents feel about the grades you get in school?

13) Do you think most of the rules in this school are fair?

23) I feel as if I really belong in this school.

25) Everything considered, are you very happy, usually happy or not too happy these days?

27) In the morning are you usually glad to go to school?

29) Do you usually like school?

39) Teachers like me.

Factor 2 (Self-rated ability)

21) I feel I have much to be proud of.

30) I am good at my work.

34) I am good in sports.

35) I am good at drawing things.

37) I am good in music.

38) I am good at making things with my hands.

Factor 3 (Internality)

12) At school, are you often blamed for things that just aren't your fault?

18) When bad things are going to happen, they just are going to happen no matter what you try to do to stop them.

20) Good luck is just as important for success as hard work.

22) Some kids are just naturally lucky.

28) When you get punished at school, does it usually seem it's for no good reason at all?